

# THE EFFECTS OF SCHOOL REFORM UNDER NCLB WAIVERS: EVIDENCE FROM FOCUS SCHOOLS IN KENTUCKY

## **Sade Bonilla**

(corresponding author)  
College of Education  
University of Massachusetts  
Amherst  
Amherst, MA 01003  
sadebonilla@umass.edu

## **Thomas S. Dee**

Center for Education Policy  
Analysis  
Stanford University  
Stanford, CA 94305  
tdee@stanford.edu

## **Abstract**

Under waivers to the No Child Left Behind Act, the federal government required states to identify schools where targeted subgroups of students have the lowest achievement and to implement reforms in these “Focus Schools.” In this study, we examine the Focus School reforms in the state of Kentucky. The reforms in this state are uniquely interesting for several reasons. One is that the state developed unusually explicit guidance for Focus Schools centered on a comprehensive school-planning process. Second, the state identified Focus Schools using a “super subgroup” measure that combined traditionally low-performing subgroups into an umbrella group. This design feature may have catalyzed broader whole-school reforms and attenuated the incentives to target reform efforts narrowly. Using regression discontinuity designs, we find that these reforms led to substantial improvements in school performance, raising math proficiency rates by 17 percent and reading proficiency rates by 9 percent.

[https://doi.org/10.1162/edfp\\_a\\_00275](https://doi.org/10.1162/edfp_a_00275)

© 2018 Association for Education Finance and Policy

## 1. INTRODUCTION

In the United States, the last fifteen years have marked an era of exceptional federal activism focused on the performance of public schools, particularly with regard to historically underserved students. This period began with the No Child Left Behind Act (NCLB), which brought test-based school accountability to scale in the United States and highlighted stark and enduring achievement gaps by race, ethnicity, and income. NCLB also set a 2014 deadline for all students to achieve proficiency. The broad expectation was that NCLB would be redesigned and reauthorized in advance of that deadline. However, because partisan gridlock in Congress stalled this reauthorization, U.S. Secretary of Education Arne Duncan articulated, in 2011, conditions under which states could receive NCLB waivers. These waivers reflected the Obama Administration's vision for the reauthorization of NCLB but proved controversial, generating pushback from both parties in Congress, as well as teachers' unions and state education agencies (Klein 2015). The goals of these waivers included increased flexibility in school accountability systems, a focus on college and career-ready standards, and new systems for evaluating the effectiveness of teachers and principals.

However, a defining feature of these waiver-driven reforms was also the emphasis on "differentiated accountability" for schools. In particular, states that received waivers were required to categorize the 10 percent of their schools with the largest achievement gaps as "Focus Schools" and the 5 percent of schools with persistently low achievement as "Priority Schools." Both reforms were effectively unfunded but otherwise differed in interesting ways. Like schools that had recently received federal School Improvement Grants, Priority Schools were required to implement a school-turnaround model consistent with federal guidance.<sup>1</sup> In sharp contrast, Focus Schools were afforded substantial flexibility as to how they identified and implemented reforms intended to reduce their achievement gaps. Understanding the effects of these federally-catalyzed reforms is particularly salient as these approaches to school accountability continue under the recent reauthorization of NCLB known as the Every Student Succeeds Act (Klein 2016).

In this study, we examine the implementation and impact of Focus School reforms in the state of Kentucky.<sup>2</sup> Kentucky is a particularly interesting setting in which to examine the effects of these reforms for two broad reasons. One is that, during this period, Kentucky established a reputation for its early adoption and high-fidelity implementation of new reforms (e.g., the Common Core State Standards, Race to the Top). Therefore, Kentucky is likely to constitute a strong test of whether these federally motivated reforms can drive meaningful improvements in school-level outcomes. Interestingly, Kentucky chose a relatively prescriptive approach to its Focus Schools that included an emphasis on school improvement planning and teacher professional development. Second, along with a few other states, Kentucky implemented a unique design feature. Kentucky identified Focus Schools using the performance of a "super subgroup" of traditionally low-performing subgroups (i.e., free or reduced-price lunch-eligible

1. See Dee (2012) for an overview of federal turnaround of these models (i.e., transformation, turnaround, restart, and closure) and their features (e.g., hiring a new principal, replacing a large portion of teaching staff, implementing instructional reforms and socioemotional supports) and their effects.

2. We do not examine the effects of the Priority School reforms because the small number of Priority Schools (and their overlap with earlier School Improvement Grant reforms) implies attenuated statistical power.

students; students with disabilities; students who were black, Hispanic, American Indian, and limited English-proficient [LEP]) for two reasons. Firstly, this approach increased the number of smaller rural schools held accountable under the new system.<sup>3</sup> Secondly, this inclusive design ensured that small subgroups that might not meet size requirements within a school would receive attention. This may be relevant given that accountability systems targeting narrow subgroups can induce schools to engage in strategic behaviors like targeting efforts among specific populations of students (Figlio and Loeb 2011). Recent and concurrent work in other states on Focus School reforms show no or potentially negative effects on student performance (Dee and Dizon-Ross 2017; Dougherty and Weiner 2017; Hemelt and Jacob 2017). Kentucky's unique omnibus approach to identifying Focus Schools could sharply attenuate the incentives for such strategic behaviors.

We rely on regression discontinuity (RD) designs to identify the causal effects of these reforms by using the discontinuous assignment of schools to Focus School status. Our RD results indicate that Kentucky's Focus School reforms led to substantial gains in students' math and reading proficiency rates (i.e., a 17 percent improvement in mathematics proficiency and 9 percent improvement in reading proficiency). Furthermore, we find evidence that these gains occurred throughout the distributions of math and reading achievement (i.e., not just proximate to the proficiency threshold). We also find that these results are robust to a variety of checks for internal validity, including possible confounds related to the nonrandom, post-reform sorting of students. We present evidence suggesting that comprehensive school planning and effective teacher professional development were important mediators of the Focus School reform's impact. Our study is organized as follows. In section 2, we discuss the policy context in Kentucky and provide a brief review of the relevant literature and theoretical considerations. In section 3, we characterize the treatment contrast for Focus Schools. In section 4, we describe the data and the construction of our analytic samples. In section 5, we present our RD estimation strategy and related robustness checks. We discuss our results in section 6 and conclude in section 7.

## 2. BACKGROUND AND PRIOR LITERATURE

The Focus School reforms examined in this study were implemented under the recent federal waivers to the NCLB Act. NCLB was the influential and controversial federal legislation that brought test-based school accountability to scale in the United States and required all students to be proficient by 2014. Anticipating that states would not be able to meet this proficiency requirement, the U.S. Department of Education began offering waivers in 2011 to states that developed more flexible systems of school accountability consistent with new guidelines. Among other things, these waivers mandated that states identify and target reforms in schools that contributed the most to achievement gaps (i.e., "Focus Schools").<sup>4</sup> Learning from these early Focus School reforms is highly relevant, as they will continue nationwide under NCLB's recent reauthorization

---

3. In Kentucky, subgroups with at least twenty-five students count for accountability purposes. By combining traditionally low-performing subgroups into an umbrella group, more schools receive an accountability rating.

4. We describe the federal NCLB waivers and Focus School reforms in more detail in section 3.

as the Every Student Succeeds Act. This current policy emphasis on a more targeted “differentiated” approach to school accountability represents the most recent iteration of a quarter-century of school accountability policies. In general, these school accountability reforms articulated consequential sanctions and supports (e.g., labeling, staff or school reconstitution, technical assistance) for schools that failed to meet test-based performance goals.

The implied theoretical motivations for school accountability situate the chronic underperformance of some schools with respect to information asymmetries and the collective-action challenges of organizational effectiveness. More specifically, the logic of these reforms is that the public-sector agents in districts and schools may lack knowledge of their performance (e.g., the existence and character of achievement gaps) or a willingness to focus on them instead of other goals without more oversight from key stakeholders (e.g., parents and taxpayers). Also, the external incentives created by school accountability may also encourage schools to overcome the collection action challenges that can inhibit school effectiveness (e.g., school culture, cooperative management, and instructional practices).

The empirical literature on school accountability generally suggests that these reforms generated meaningful, though not transformational, improvements in student learning, as measured by low-stakes tests unrelated to the accountability incentives (Figlio and Loeb 2011). This quasi-experimental literature includes studies of the state-level reforms that preceded NCLB (Carnoy and Loeb 2002) as well as studies of NCLB’s impact (Dee and Jacob 2011; Wong, Steiner, and Cook 2013). There is also evidence that schools and teachers sometimes respond to the presence of accountability incentives (see Figlio and Loeb [2011] for a summary). For example, Dee, Jacob, and Schwartz (2013) find evidence that NCLB led to a modest reallocation of instructional time from non-tested to tested subjects.

The Focus School reforms are also closely related to “whole-school” or comprehensive school reforms (CSR) that call for the simultaneous and multifaceted overhaul of an entire school to refocus efforts to raise student achievement (Borman et al. 2003). Between 1998 and 2005, the U.S. Department of Education disseminated \$1.8 billion in grants for low-performing schools to participate in a CSR program. The federal CSR program articulated eleven broad reform strategies required of participating schools. These included a focus on strong academic content, articulating measurable goals, professional development, attention to the allocation of resources, and outside technical assistance. Studies of the federal CSR initiative found little evidence that they were successful in raising student achievement (Gross, Booker, and Goldhaber 2009; USDOE 2010). Other evidence suggests that the school-level implementation of the federal CSR model was uneven in ways that mattered for sustaining school improvement (Desimone 2002; Bifulco, Duncombe, and Yinger 2005; USDOE 2010). A meta-analysis by Borman et al. (2003) concludes that, in contrast to the federal CSR effort, other specific CSR models (i.e., Direct Instruction, Success for All, and the School Development Program) appear to have had meaningful impacts on student achievement, particularly when implemented over several years.<sup>5</sup>

5. Focus School reform efforts are distinct from school turnaround models described in Dee (2012), Papay and Hannon (2015), Sun, Penner, and Loeb (2017), and others. School turnaround reforms were characterized

Recent research on the NCLB waiver accountability reforms has been almost uniformly negative. In Michigan, flat performance of higher-achieving students drove improvements in within-school achievement gaps (Hemelt and Jacob 2017). Similarly to our study, Focus School status does not appear to influence staffing, student composition, or enrollment patterns, however, Hemelt and Jacob report uneven implementation of the reforms. This stands in stark contrast to the unique setting in Kentucky that relied on capacity built prior to the issuance of NCLB waivers. In Rhode Island, researchers find mixed results but rely on a small sample size of nine treated Focus Schools (Dougherty and Weiner 2017). In a better-powered study in Louisiana, Dee and Dizon-Ross (2017) find no impacts on student achievement at Focus Schools under a system that also assigned letter-grades to schools. In our study, we leverage data from a high-fidelity implementation state that carried out a unique design that may have discouraged targeting students and that provides rich contextual data on potential candidate mediators.

These studies have several implications for the Focus School reforms studied here. First, the prior evidence suggests some optimism that the latest incarnation of consequential school accountability (i.e., targeting particular schools for improvement and supports) may be effective in improving student achievement. Second, this literature also implies that attention to the potential unintended consequences of such reforms is warranted. Third, the evidence from successful CSR models suggests Focus School reforms may be more effective to the extent they use similar strategies, such as formative assessment and data-driven instruction (e.g., Success for All), schoolwide planning and community engagement (e.g., the School Development Program), and differentiated instruction (e.g., Direct Instruction). Finally, the experience with the federal CSR program implies the efficacy of Focus School reforms will turn critically on the quality of the implementation at the state and local levels. In the next section, we turn to such evidence by characterizing Kentucky's Focus School reforms, as well as the evidence on their implementation.

### 3. FOCUS SCHOOLS IN KENTUCKY

For the past twenty-five years, Kentucky has situated itself as a consistently early and enthusiastic adopter of education reforms. For example, in 1990, the Kentucky Education Reform Act reformed the state's education finance system, increased school accountability, and introduced new educational standards and assessments. More recently, the state was the first to adopt the Common Core State Standards (CCSS), doing so before the standards were complete (Governor's Task Force 2011; Ujifusa 2013). Kentucky was also viewed as a leader in supporting the classroom implementation of the CCSS (e.g., Gewertz 2011). The state received \$17 million in the federal Race to the Top competition and was among the first group to apply for a federal waiver from NCLB. Furthermore, the U.S. Department of Education later invited Kentucky to apply for a "fast-track" renewal of its NCLB waiver (Klein 2015). Kentucky's recent record of adopting and implementing leading K–12 policy reforms suggests that it is a particularly propitious setting in which to study the effects of the school reforms required under NCLB waivers.

---

by large multiyear grants and requirements to significantly restructure schools (leadership or staff changes, closure, etc.).

### NCLB Waivers

The waiver process had its genesis in both the design of NCLB and federal politics. NCLB required that all students be deemed proficient in their state accountability system by 2014. The widespread anticipation that schools would fall short of this ambitious goal motivated extensive discussion of reauthorizing NCLB with updated requirements. However, the legislative failure to advance such legislation created the opportunity for the U.S. Department of Education to offer states waivers from these requirements. In 2011, under Secretary Arne Duncan, states were encouraged to apply for such waivers by outlining new accountability systems consistent with the reform principles favored by the Obama Administration. Central among these was the encouragement to adopt “differentiated recognition, accountability, and support systems.” States receiving waivers were required to identify schools that were persistently lowest performing (i.e., Priority Schools) and have them implement federally prescribed reforms.<sup>6</sup>

They were also required to designate as Focus Schools those that had the lowest performance for designated subgroups or the largest performance gaps between subgroups. A minimum of 10 percent of Title I schools were required to be identified as Focus Schools. States were required to field interventions in these schools, although the federal waivers allowed flexibility in contrast to the prescriptive requirements associated with Priority Schools. Whereas Priority Schools were required to implement one of the federal School Improvement Grant turnaround strategies, Focus Schools were asked to implement interventions that were “consistent with” federal turnaround principles or any other “research-based” interventions to meet the needs of students at the school (USDOE 2012). We describe Kentucky’s Focus School initiative in more detail below.

Kentucky submitted its application for an NCLB waiver in November 2011. The U.S. Department of Education approved the waiver in February 2012 and implementation was set to begin in the 2012–13 school year. However, a vendor delay slowed implementation of the identification of schools and, as a result, the Kentucky Department of Education (KDE) notified schools of their Focus School status in late October of 2012. Schools then had 90 days to develop their Comprehensive School Improvement Plan (CSIP), a central element of the waiver and Focus School reforms that we describe in more detail below. Given the delayed timeline, schools did not have a full school year to implement reforms. Therefore, in this paper we focus our analyses on outcomes from the 2013–14 school year, the first full school year in which Focus Schools implemented their chosen reforms.

We later describe in detail Kentucky’s procedure for identifying Focus Schools. However, one distinctive feature merits special attention. As a predominantly rural state, Kentucky has many small schools that often do not enroll enough students to meet the minimum subgroup size.<sup>7</sup> In order to have appropriate representation of

6. The number of Priority Schools is required to be equal to 5 percent of the number of Title I schools in the state. We do not study these reforms in Kentucky because there are too few of such schools to support reasonable statistical power.

7. The minimum subgroup size in Kentucky is twenty-five students. A subgroup that meets the minimum subgroup threshold has their achievement factored into a school’s accountability rating and reports their achievement publicly.

schools serving targeted students among designated Focus Schools, Kentucky introduced a “super-subgroup” measure, combining unduplicated counts of students from traditionally low-performing subgroups (i.e., free or reduced-price lunch–eligible; students with disabilities; students identifying as black, Hispanic, American Indian, and LEP) into an umbrella group. To mitigate any concerns that the larger group would mask low individual subgroup achievement, Kentucky also used a second method for Focus School identification that considered subgroup performance of traditionally low-performing subgroups separately. A seemingly unintended but possibly beneficial feature of the “super-subgroup” measure is that it attenuates any incentives schools may have to focus on a particular subgroup of students, possibly at the expense of others (i.e., “triage” behavior). The “super-subgroup” identification strategy, in various forms, was also used by sixteen additional states in their NCLB waivers plans and continues to be utilized by states under the Every Student Succeeds Act (Polikoff et al. 2014; Klein 2017).

### **Focus School Reforms in Kentucky**

Kentucky’s leading transition to the CCSS also provided the state with motivation to redesign its school accountability system. In 2009, bipartisan legislation (Senate Bill 1) called for an overhaul of previous reform efforts and a unified focus on college and career-ready standards and assessments to be implemented in the 2011–12 school year. In its 2011 waiver application, the state noted the strong alignment between these in-progress reforms, the corresponding changes to their accountability system, and the principles behind NCLB waivers. Kentucky’s newly designed accountability system focused on holding all schools accountable for student achievement.<sup>8</sup> In particular, the central tenet of Kentucky’s accountability system (USDOE 2011) focused on the design, revision, and monitoring of CSIPs. A school committee that includes representation from parents, students, and the community would begin the state-designed CSIP process with a school needs assessment linked to data, explicit standards, and indicators for school improvement. The CSIP guides these committees to identify and prioritize their school’s needs. They then identify strategies and activities designed to address the high-priority needs, a person responsible for that task, a corresponding timeline, and a funding source.

The CSIP process is critical for understanding the treatment contrast created by Focus School eligibility. In particular, as articulated in Kentucky’s original waiver application, designation as a Focus School implied several additional CSIP requirements related to gap issues and the school’s plan to address them. For example, districts were required to assist their Focus Schools with their needs assessment, using guidance from the Commissioner’s Raising Achievement/Closing Gaps Council. The CSIPs for Focus Schools also had a unique requirement to identify specific strategies for closing gaps in student achievement and graduation rates. The state also required the Focus School CSIPs to discuss curricular alignment that would ensure a rigorous instructional program aligned to student needs and the Common Core. Other unique requirements of Focus School CSIPs consistent with “whole school” strategies were the

---

8. As such, all schools—including those ineligible for Title I support—were eligible for Focus School status if they met the selection criteria.



provision of time for collaborating on the use of data to inform instructional strategies, aligned professional development, and strategies for school safety and discipline. All schools were required to post their CSIPs on their Web sites so they would be available for public inspection.

The KDE also established units dedicated to supporting school capacity to implement these plans. The Office of Next Generation Schools and Districts orchestrated technical assistance to schools statewide. However, a specialized KDE unit, District 180, was specifically tasked with providing Focus and Priority Schools technical assistance in developing and implementing their CSIPs (e.g., identifying research-based interventions). The state also sought to publicize school improvement efforts. The KDE encouraged Focus Schools to submit potential best practices, which they graded according to a standard rubric and posted these vetted practices in an online database for other schools to access. In terms of monitoring Focus School improvement, the KDE viewed local districts as the first level of oversight while KDE staff provided additional oversight when schools were not on track to exit Focus School status (KDE 2011). Focus Schools that failed to exit Focus School status after several years could also be redesignated as Priority Schools and compelled to undertake more explicit reforms.

Kentucky's original waiver application explicitly indicates that the CSIP requirements unique to Focus Schools are the defining characteristic of their intended intervention. However, apart from the CSIP process, there were at least two other candidate mediators that may characterize the Focus School intervention. One is simply the stigma that may be associated with the Focus School label and that could spur school improvement independent of the CSIP process. Another is that Kentucky's waiver reforms provided Focus Schools with enhanced financial flexibility to combine and redirect multiple Title I funding streams. It also expanded eligibility for the Title I Schoolwide Program (SWP) to all Focus Schools.<sup>9</sup>

After the first full year of Focus School implementation (i.e., in August 2014), Kentucky submitted its successful application for an extension to its NCLB waiver. This application indicates that a key treatment contrast between Focus Schools and other schools was attenuated by the 2014–15 school year. That is, the application describes a reinvigorated emphasis on supporting and monitoring CSIPs through a new review rubric that applies to all Focus and Title I schools.<sup>10</sup> Because of this extension of the key intervention elements to all Title I schools (i.e., the overwhelming majority of our analytical sample), we focus our analysis on outcomes from the 2013–14 school year (i.e., the first full year of Focus School implementation) in which the treatment contrast was crisp.<sup>11</sup> Nonetheless, we also report the results from the 2014–15 school year in a separate online appendix that can be accessed on *Education Finance and Policy's* Web site at [www.mitpressjournals.org/doi/suppl/10.1162/edfp\\_a\\_00275](http://www.mitpressjournals.org/doi/suppl/10.1162/edfp_a_00275).<sup>12</sup>

9. As a practical matter, this funding flexibility has attenuated empirical relevance for the treatment contrast we study. Specifically, among the 187 Focus Schools in our analytical sample, only 42 obtained SWP status because of the waiver reforms.
10. We examined this rubric and found that it contained the CSIP elements that had been required only of Focus Schools through the 2013–14 school year.
11. The treatment contrast may have also been attenuated in subsequent years by a growing awareness of the threat of Focus School status among schools above but close to the eligibility threshold.
12. These results in online appendix table A.1 suggest similarly positive effects of the Focus School status. However, the point estimates are somewhat smaller and statistically insignificant. This pattern is consistent with the



### Focus School Implementation

Although our core contribution comes from our credibly causal estimates of the Focus School classification, the previous literature suggests that implementation fidelity is linked to outcomes. As such, in this next section we characterize the implementation of Kentucky's Focus School reforms to inform our understanding of the treatment effects.

The activities unique to Focus Schools featured school improvement planning with an emphasis on state-approved strategies (e.g., gap-closing programs and activities, instructional reforms, and professional development) coupled with local and state monitoring as well as financial flexibility. However, the actual implementation of these reforms may have differed from the stated policy intent. We look to several sources of information, including federal monitoring reports of Kentucky's waiver activities, evidence from media accounts, our own interviews with KDE officials responsible for supporting Focus Schools, our own review of school improvement plans, and results from a statewide teacher survey.

A federal review of Kentucky's waiver reforms occurred in August 2013 (i.e., just before the first full year of Focus School implementation). Federal officials reviewed state documents and conducted interviews with several state officials before issuing annual monitoring reports summarizing implementation and listing next steps when implementation did not comply with the approved waiver. In this report, the federal review called for the state to increase their oversight and monitoring in concordance with their approved waiver (USDOE 2013). Kentucky's successful 2014 application for a waiver extension indicated that the U.S. Department of Education was satisfied the state addressed the areas of concern from the monitoring report (Delisle 2014).

We now turn to implementation evidence about the experience of Focus Schools from local media coverage that includes interviews with school principals as well as our personal communications with KDE officials. Media coverage of Focus Schools in Fayette County Public Schools (FCPS), the second largest school district in the state, indicates that Focus School principals were required to present their school improvement plans to the school board. District officials also assigned principals to work with a mentor, usually a retired principal, to help implement the plans (Spears 2015). Other media coverage indicates that staff and the local community were aware of the label and highly motivated to exit Focus School status.

Personal communication with state officials on technical assistance and state monitoring revealed that Focus Schools may not have received the high level of state support and oversight articulated in the approved waiver (i.e., a finding also suggested by the federal monitoring report from 2013). A KDE official disclosed to us that the state office charged with providing technical assistance primarily served Priority Schools, although services were accessible to Focus Schools. We also learned that the staff at the KDE technical assistance office, District 180, interacted with Focus Schools mainly by reviewing their CSIP. Unfortunately, no records of the frequency or intensity of these interactions exist. Although the state articulated an exhaustive protocol for the monitoring and oversight of Focus Schools, the onus appears to have been on local districts to provide oversight and guidance to treatment schools.

---

relevance of the school planning process as a mediator and the clear dampening of the treatment contrast described here. However, we acknowledge the possibility that the intervention's effects faded.

With this in mind, we examined school improvement plans to gain additional insight into the character of the Focus School activities. We conducted a review of CSIPs for all schools in the two largest school districts in the state, Jefferson County Public Schools (JCPS) and FCPS that had plans readily available for review. We obtained CSIPs for all one hundred seven JCPS and forty-seven FCPS schools in our sample for the 2013–14 school year.<sup>13</sup> In our review, we coded whether schools mentioned the major whole-school reforms recommended by the state (e.g., extended learning time, teacher collaboration, professional development, data-driven instructional practices, and computer-aided instruction for low-performing students) and find nearly universal compliance. However, schools were somewhat less likely to follow KDE guidance in listing specific strategies for students with disabilities and/or non-English proficient students (i.e., 61 percent). Our review of CSIP plans suggests that Focus Schools followed the recommended state guidelines in developing their plans but this does little to confirm whether Focus Schools carried out their plans differently from those of non-treatment schools.

To this end, we turn to our final source of information from state surveys of teachers that provides information on the experience of teachers rather than intended activities. The Teaching, Empowering, Leading and Learning Kentucky survey, administered in 2015, covered salient topics on teacher working conditions including professional development, use of time, instructional practices, and support. The high average response rate (92 percent) allows us to examine whether teachers at Focus Schools report differences compared with those at nontreatment sites. Teacher responses to the survey indicate that the quality, rather than the quantity, of professional development sessions may explain the performance gains. We discuss these results in more detail in the Results section. In summary, the available evidence suggests that Focus Schools followed state guidance in developing key operational details of their plans for school improvement.

#### 4. DATA AND SAMPLE CONSTRUCTION

We use cross-sections of school-level data with information on relevant subgroup performance to examine the effect of the Focus School reform. Our data come from the annual School Report Cards (SRC) published by the KDE over several years. Our core outcomes are school-level measures of student test performance (e.g., percent proficient) by subgroup for the first full year of implementation (2013–14). The SRC also provides school-year data on the baseline assignment variables, Focus School status, and other characteristics of schools, students, and teachers. We also add school level data from the Common Core of Data on eligibility for and participation in federal Title I programs, as well as school directory information.

To construct the analytic sample, we include all schools in the risk pool for Focus School status following the multiple rating assignment procedure. KDE provided 1,296 schools with a differentiated accountability rating but we restrict the analytic sample in several important ways. First, we exclude all 230 high schools because the end-of-course

13. Nearly half of JCPS middle and elementary schools were classified as Focus Schools ( $N = 51$ ). FCPS have seventeen Focus Schools in the sample. Together, these two large districts host sixty-eight Focus Schools, or 36 percent, of the Focus Schools in the sample.

exams administered to these students are not part of the state's Kentucky Performance Rating of Education Progress (K-PREP) testing system and separate analyses of high schools are uninformative due to insufficient power. Second, we exclude 12 schools for one of three reasons: (1) The state identified nine of these schools as Priority Schools, implying ineligibility for Focus School status;<sup>14</sup> (2) We eliminate a single school because it did not meet the minimum student group size to receive an assignment score;<sup>15</sup> (3) We also exclude two schools that are part of a PK–12 laboratory school situated on Eastern Kentucky University's campus that require an application for admission and charge tuition and fees. These edits resulted in a sample of 1,054 elementary and middle schools eligible for Focus School status.

Our sample is limited further by the exclusion of schools that have a valid assignment variable but lack outcome data (i.e., test scores). Schools lack valid outcome data for one of two reasons: school closure ( $N = 27$ ) or a grade reconfiguration such that the school no longer has assessment results for the content level in which it received its original accountability assignment ( $N = 11$ ). The exclusion of these 38 schools, which are located in small, rural districts, reduces our sample further to 1,016 schools. Though these school changes are uncommon, we find that they are concentrated among schools that were not designated as Focus Schools (see online appendix figure A.1). In fact, parametric RD estimates indicate that Focus School status significantly reduced the probability of school closure or reconstitution.<sup>16</sup> Given the policy debates around school closures, we view this as a substantively interesting finding (i.e., that an energetic school reform effort reduced the risk of a school's being closed or reorganized).

However, in terms of evaluating the impact of Focus School reforms, these school changes, though uncommon, may still constitute an internal-validity threat. Specifically, it is reasonable to suppose that the schools that were closed or reorganized would have had comparatively poor student outcomes if they had remained structured as they were. The systematic removal of such struggling schools from the "control" side of the threshold that determined Focus School status would then imply an upward bias in the estimated effects of these reforms. To address this concern, we identified the eighteen small, rural school districts in which these school changes occurred and eliminated all of the schools from these districts. This implied an exclusion of only ninety-six additional schools and reduced our final analytical sample to 920 schools.<sup>17</sup> We focus on this sample because of the clear integrity of the intent-to-treat population. However, we also note that the larger, inclusive sample implies similar results, suggesting the

14. Kentucky identified forty-one Priority Schools (thirty-two high schools and nine middle schools). No elementary schools were given Priority School status.

15. Anchorage Independent Public School is a K-8 school that is the sole school in the school district. The elementary level classification was "High performing" while the state classified the middle school as a reward school. The school did not have enough students in the gap group (or any of the individual subgroups) to present results. They are the only school in Kentucky that did not report information on gap group students at baseline.

16. We look both parametrically and nonparametrically for discontinuities at the assignment threshold for school closure and grade reconstitution (both combined and separately). Our estimates confirm the statistical significance of the discontinuities.

17. We adopted a parallel sample-construction protocol for the 2014–15 results (available in the online appendix). This results in a sample of 852 schools.

**Table 1.** Descriptive Statistics of Regression Discontinuity Samples

Variable	Mean	Standard Deviation	Minimum	Maximum
<b>Accountability characteristics</b>				
Focus School 2013–14	0.203	0.403	0	1.0
SGG score (centered)	12.5	9.8	–19.8	55.0
I(SGG Score < 0)	0.092	0.290	0	1.0
Reading proficiency rate, gap group 2013–14	46.2	10.0	17.2	85.7
Math proficiency rate, gap group 2013–14	38.8	11.5	0.0	78.6
<b>Student characteristics</b>				
Male	0.512	0.035	0	0.619
White	0.827	0.197	0.073	1.0
Black	0.093	0.144	0	0.767
Hispanic	0.043	0.061	0	0.739
Asian	0.012	0.021	0	0.218
Free/reduced-price lunch	0.630	0.200	0.028	1.0
Attendance	0.952	0.012	0.892	0.979
<b>School characteristics</b>				
Elementary	0.704	0.457	0	1.0
Middle	0.296	0.457	0	1.0
Enrollment	475	195	79	1,625
Student–teacher ratio	15.5	2.2	10	21
Rural	0.500	0.500	0	1.0
Urban	0.173	0.378	0	1.0
Suburban	0.134	0.341	0	1.0
Reward school	0.053	0.225	0	1.0
Title I eligible	0.919	0.274	0	1.0
Title I Schoolwide program	0.799	0.401	0	1.0
<b>Teacher characteristics</b>				
Teaching experience (years)	11.7	2.5	2.3	20
Advanced degree (Masters, Ph.D.)	0.513	0.126	0.071	0.917

*Notes:* The sample includes 920 schools with an SGG (Student Gap Group) assignment variable. The SGG Score refers to the baseline assignment variable based on 2011–12 test scores and is centered on the eligibility threshold. The student–teacher ratio is recoded (winsorized) at the 1st and 99th percentiles.

*Sources:* Student, school, and teacher characteristics are authors' calculations using Kentucky School Report Card and Common Core of Data, 2011–12.

external-validity caveats associated with deleting these rural districts is not empirically consequential.<sup>18</sup>

In table 1, we present descriptive statistics for this sample of 920 schools. Our core outcome measures are the reading and math proficiency rates for the targeted umbrella “Gap Group” in 2013–14. In the publicly reported data, there is some modest suppression of these test score results because of privacy concerns related to small cell sizes and the restrictions of the Family Educational Rights and Privacy Act. For our core test-score outcomes, only two schools suppress their reading performance and four schools

18. We show these results in our online appendix (table A.1). The first column shows results from our preferred sample (i.e., including only schools in districts without school closures or grade configuration changes). The second column includes all schools with available results and, finally, the third column uses the Last Observation Carry Forward (LOCF) imputation procedure. LOCF carries forward the last available proficiency rate for schools in the sample with missing outcomes.

suppress their math performance. Given how uncommon this suppression is (and the procedure for being allowed to do so), there is little reason to suspect that it constitutes an internal-validity threat. Nonetheless, for every test-score outcome and specification (e.g., fuzzy and frontier RD) we report, we examined auxiliary RD models that examined whether test-score suppression was balanced around the threshold. We consistently found no evidence that suppression changed significantly at the Focus School threshold.

Kentucky's SRC contain a rich set of covariates at the school level, including content level (e.g., elementary, middle, high), total enrollment, student-teacher ratio, attendance, and teacher qualifications (table 1). We supplement SRC data with additional measures from the Common Core of Data, including school urbanicity, Title I eligibility, and Title I program type. We utilize performance level assessment data from the K-PREP. These K-PREP assessment data are aligned with the CCSS and are available in mathematics and reading for grades 3–8 aggregated at the school level for different subgroups. Performance data report the percent of students achieving a particular performance level: novice, apprentice, proficient, and distinguished. Although we focus on the proficiency rates of “gap group” students (i.e., students identifying as free or reduced price lunch-eligible, special education, black, Hispanic, American Indian, and LEP), we also explore treatment heterogeneity at different performance thresholds and for the available subgroups. The 2011–12 school year was the first year of the K-PREP assessment program (i.e., the first state to administer CCSS-aligned assessments). The identification of Focus Schools under Kentucky's NCLB waiver, which we describe in more detail below, is from this first year of K-PREP assessments.

Table 1 lists baseline characteristics for the 920 schools in the analytic sample that are eligible for Focus School status. About one-fifth of the analytic sample, 187 schools, have Focus School status because they meet one or both of the criteria for identification. The Student Gap Group (SGG) score, based on 2011–12 K-PREP results, is one of the assignment variables that determines Focus School status. We center the SGG score so that schools with a value less than zero are eligible for Focus School status (i.e., 9.2 percent of the sample meets the SGG criteria).

Student characteristics include the percent of male students, percent of students qualifying for free/reduced-price lunch, racial/ethnic composition (white, black, Hispanic, etc.), and attendance rates. Kentucky schools largely enroll white students; the average school enrolls 82.7 percent white students, though some regions enroll relatively small shares of white students and larger proportions of black and Hispanic students. Asian, American Indian/Alaskan Native, and other groups constitute a small proportion of Kentucky students. Because of the overall low enrollment numbers of black and Hispanic students and their concentration in a minority of schools, we do not have statistical power to detect treatment effects separately for these subgroups but note that they are included in the gap group. Students in Kentucky are relatively socioeconomically disadvantaged—the average school has 63 percent of students eligible for free or reduced-price meals. Kentucky is a predominantly rural state—half of the schools in the analytic sample are located in rural areas and less than 1 in 5 are located in urban areas. Although information on LEP and special education program participation is not available at baseline, they are available for the post-treatment period and we use these data to conduct additional robustness checks.

Almost 80 percent of schools administer a Title I SWP at baseline, a designation that prior to the Elementary and Secondary Education Act waiver required schools to enroll at least 40 percent free/reduced-price lunch–eligible students.<sup>19</sup> Teacher characteristics include teachers' mean years of experience and the proportion with an advanced degree. Although the average school has a teaching staff with over ten years of experience and just over half have advanced degrees, this varies dramatically across schools. Average teacher experience ranges from two to twenty years, and the proportion of teaching staff with an advanced degree ranges from 10 to 90 percent.

## 5. REGRESSION DISCONTINUITY DESIGN AND ESTIMATION

### RD Designs

In order to obtain unbiased estimates for schools at the discontinuity, the RD identification strategy relies on the assumption that the only change at the threshold is the probability of Focus School classification. In other words, those schools who failed to achieve a predetermined level of proficiency face a requirement that schools with non-Focus School status do not. Our approach to analyzing the resulting discontinuities in treatment status relies on a relatively weak assumption: the average outcome of Focus Schools is a continuous function of the assignment score. We discuss the assignment mechanisms, our estimation strategy, and falsification tests that support the compelling causal warrant of this RD design.

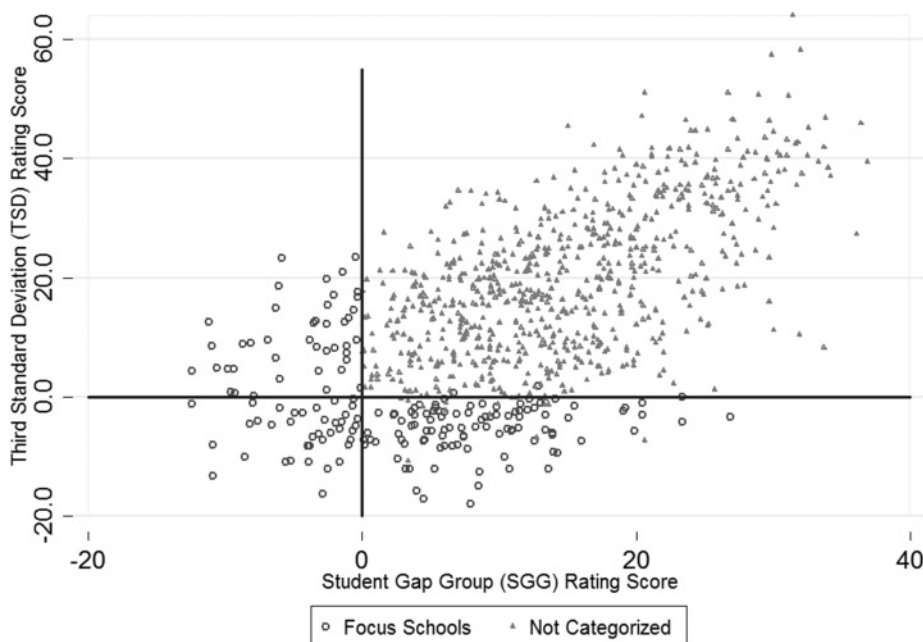
### Assignment Mechanisms

In this study, we focus on one of the two assignment criteria for Focus School classification calculated by the KDE. The SGG score is an average of the percent of proficient or above for “gap group” students across five subject areas in the 2011–12 K-PREP assessments.<sup>20</sup> Gap group students with membership in multiple subgroups are counted once in the formula. Therefore, the group is based on the unduplicated count of students with membership in any gap group (i.e., free or reduced-price lunch, special education, black, Hispanic, American Indian, and/or LEP). The threshold for Focus School status is specific to each school level—schools with an SGG score in the bottom 10 percent of elementary, middle, or high schools are eligible for Focus School classification.

A second assignment criteria, the Third Standard Deviation (TSD) method, targets schools with low subgroup performance relative to the statewide proficiency rate in any of the five designated subjects. Any school with a minimum of twenty-five tested students in a given subgroup would receive the Focus School designation if the subgroup-by-subject performance met the TSD assignment criteria. For each level (i.e., elementary, middle, and high school), KDE calculated the mean and standard deviation

19. As noted earlier, 42 of the 187 Focus Schools in our sample obtained SWP status as part of the reforms, implying that funding flexibility is part of the treatment contrast. Visual inspection of SWP status suggests a modest increase at the eligibility threshold. However, auxiliary RD regressions that condition on the control variables indicate that SWP status appears continuous through the Focus School threshold, suggesting it is not a salient part of the contrast.

20. The five subject areas used for the SGG assignment variable are reading, math, science, social studies, and writing/language mechanics. These proficiency rates of gap group students are averaged to calculate the SGG score. For our outcomes, we focus on reading and math because these are the only subjects in which students are tested annually in grades 3 through 8. Students are tested in science in grades 4 and 7; social studies in grades 5 and 8; writing in grades 5, 6, and 8; and language mechanics in grades 4 and 6.



Notes: This graph illustrates Focus School assignment along both forcing variable dimensions, Student Gap Group (SGG), and Third Standard Deviation (TSD) rating scores, on the x- and y-axis, respectively. Schools with rating scores less than zero on either dimension (i.e., quadrants II, III, and IV) meet the Focus School classification criteria.

**Figure 1.** Focus School Assignment by Forcing Variable

of the statewide proficiency rate for all students in each of the five subjects. Any school with a subgroup performing three standard deviations below the statewide mean in a subject was eligible for Focus School status. This implies that, in theory, a school could have as many as thirty TSD assignment variables (i.e., scores for six subgroups by five subjects).<sup>21</sup> However, we define each school's relevant TSD assignment variable as its lowest subgroup-by-subject performance measure with respect to the relevant TSD threshold. As a practical matter, the lowest TSD score among Focus-eligible schools was often in reading (76 percent) and the lowest-performing subgroup was, typically, special education (86 percent).

In sum, the state intended to designate a school as a Focus School if it met one of two conditions. One is when gap group students performed below the 10-percent threshold for its school level. Alternatively, a school would receive Focus status if, for any designated subject, a targeted subgroup within the school performed three standard deviations below the state mean for the group. Because schools receive a rating along both dimensions, they can fall into one of four quadrants that relate to their intent-to-treat status, as illustrated in figure 1. Quadrant I contains schools with both SGG and TSD scores above the assignment thresholds ( $N = 720$ ). These schools are ineligible for Focus School status (i.e., the intent to treat is zero). The 200 schools spread across the

21. This was not true in practice because the three-standard-deviation threshold was below zero in two subjects (math and writing). Therefore, the maximum number of TSD assignment variables per school was eighteen (i.e., six subgroups by three subjects).



remaining three quadrants qualify for Focus School status based on their SGG score (quadrant IV), their TSD score (quadrant II), or by having both SGG and TSD scores (quadrant III) below the assignment threshold (i.e., the intent to treat is 1).

Our analytic approach relies almost exclusively on the Focus School assignment generated by SGG scores (rather than the TSD scores) as a source of identifying variation. Our motivation for this emphasis is that the Focus School uptake generated by SGG scores is particularly interesting because it is likely to motivate a broader and more impactful reform than those motivated more narrowly by the low performance of just one subgroup and subject. In our concluding remarks, we underscore that the gap-group policy design we leverage (and the particular treatment contrast it created) is among several important issues that influence the external validity of our findings.<sup>22</sup>

Our preferred RD designs leverage the credibly quasi-random assignment to Focus School reforms generated by the proximity of a school's SGG score to the arbitrary threshold that determined treatment. Specifically, we adopt two approaches articulated in the recent literature on multivariate RD designs (e.g., Reardon and Robinson 2012; Wong, Steiner, and Cook 2013). One is a straightforward “fuzzy” RD design based on the full sample. In this sample, all schools with SGG scores below the threshold are subject to Focus School reforms (quadrants III and IV of figure 1). However, as we show below, a fraction of schools with SGG scores above the threshold are Focus Schools because they are assigned under the TSD rule (schools in quadrant II; hence, the “fuzziness” in the first stage assignment). This approach is akin to comparing the schools on either side of the vertical axis in figure 1 (i.e., quadrants III and IV compared with quadrants I and II). Our second approach is a “frontier” RD design based on eliminating schools that were already treatment-eligible under the TSD rule (i.e., omitting schools in quadrants II and III). This approach gives us a virtually “sharp” first-stage relationship focusing on schools that were only eligible because their SGG score fell below the threshold.<sup>23</sup> We report both results because each has potentially compelling properties. The fuzzy RD approach leverages the statistical power of the full sample. However, the frontier RD approach does not rely on the assumption that potential outcomes are continuous along a longer frontier characterized by potentially heterogeneous treatments (Wong, Steiner, and Cook 2013).<sup>24</sup> We also examined a “binding score” RD approach in which both assignment variables were converted to a single variable (i.e., the minimum of both). The binding score approach compares schools in quadrants II, III, and IV to those untreated schools in quadrant I. We find that this approach generates results similar to those we report but believe its implicit assumption (namely, that of homogeneous

22. For completeness, we also show results based on the TSD assignment variable we constructed (table A.2 in the online appendix), where we find similar school-level and subgroup improvements in students' reading performance, the subject most often targeted by this narrow assignment rule. Interestingly, we find no such improvements in math (i.e., the non-targeted subject) suggesting schools responded to how they were identified.

23. The frontier RD is not a fully sharp design because two ineligible schools mistakenly received Focus School status.

24. We refer to treatment with two issues in mind. Whereas the explicit Focus School treatment required schools to reduce the achievement gap through the strategies detailed in section 3 (i.e., treatment contrast from the NCLB waiver), the targeted students varied depending on how schools were identified (i.e., treatment contrast based on the SGG and TSD rating).

treatments along the entire response surface) is untenable given the more targeted treatment we expect in schools eligible only under the TSD rule.<sup>25</sup>

### Estimation

In this study, we report reduced-form estimates of the effect of Focus School eligibility (i.e., the intent to treat) based on specifications with the following general form:

$$Y_i = \delta I(SGG\ Score_i < 0) + f(SGG\ Score_i) + X_i + \varepsilon_i. \quad (1)$$

In this specification,  $I(SGG\ Score_i < 0)$  is an indicator for whether the centered SGG score for school  $i$  is below the threshold that implies Focus School eligibility. The variable,  $X_i$ , represents school-level baseline covariates, and  $\varepsilon_i$  is the error term. The parameter of interest,  $\delta$ , represents the level change in student proficiency at the Focus School assignment threshold conditional on a flexible function of the assignment variable,  $f(SGG\ Score_i)$ . Our preferred models impose a linear relationship between the SGG score and outcomes but allow for different slopes on either side of the threshold. We chose our preferred functional form using both graphical evidence and the Akaike information criterion (AIC) (Cook et al. 2015). However, we also estimate outcomes using higher-order polynomials. Furthermore, we also present results based on local linear regressions using only those observations in increasingly tight bandwidths around the assignment threshold (Lee and Lemieux 2010). These bandwidths include the optimal bandwidth chosen by the procedure introduced by Imbens and Kalyanaraman (2011), which is generally in the range of  $\pm 7$  to 9 rating score points. We also present estimates based on triangular kernel weights in addition to unweighted (i.e., rectangular kernel) results.

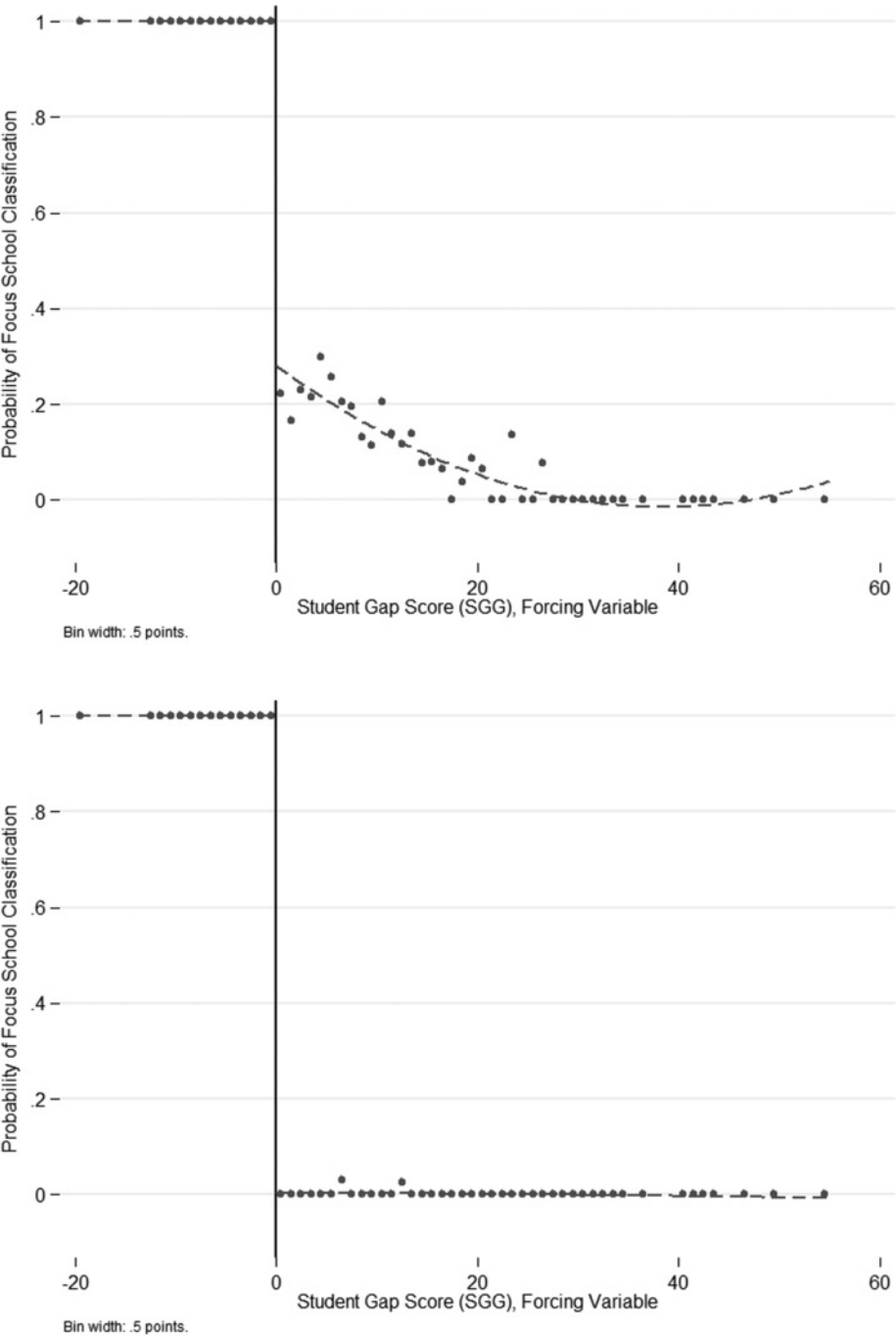
### Treatment Assignment

Before turning to our main results, we first discuss the first-stage relationship between the assignment variable and schools' participation in Focus School reforms, as well as the causal warrant of our RD designs. We begin by illustrating, both graphically and parametrically, the relationship between a school's SGG score and its participation in Focus School reforms. In figure 2a, we show the change in probability of Focus School status for the SGG assignment variable for the Fuzzy RD sample. All schools with an SGG score in the bottom 10 percent are classified as Focus Schools, but because of the secondary assignment mechanism (i.e., TSD score) the probability drops approximately 70 percentage points to the right of the threshold where only 30 percent of schools are Focus Schools. In figure 2b, the frontier RD changes in probability of treatment from one to nearly zero, a virtually sharp contrast.<sup>26</sup>

In table 2 we show the first-stage RD estimates based on the SGG assignment variable for the fuzzy and frontier RD. In the sparsest specification (i.e., column 1), the RD estimates indicate that the probability of becoming a Focus School jumps by

25. Although the regulatory designation is the same, it is likely that a school identified as a Focus School through the TSD mechanism may be more likely to "triage" their efforts narrowly to the relevant subgroup and subject.

26. Two schools that are classified as Focus Schools but do not fall below the assignment threshold for either criteria are included in the analytic sample. Results were similar when we removed these schools from the analytic sample. After communicating with KDE officials, we were unable to uncover the rationale for Focus School classification for these two schools.



Notes: Graphs of Focus School treatment status by the SGG assignment variables for the Fuzzy and Frontier RD samples. Bin width 0.5, full sample.

Figure 2. Focus School Status by Student Gap Group (SGG) Score, 2013–14. (a) Fuzzy Regression Discontinuity (RD). (b) Frontier RD.

**Table 2.** The Effect of Focus School Eligibility on Focus School Status

Independent Variable:	Dependent Variable: Focus School 2013–14							
	Fuzzy RD				Frontier RD			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
I(SGG Score <sub>i</sub> < 0)	0.746*** (0.025)	0.700*** (0.038)	0.698*** (0.038)	0.684*** (0.047)	0.994*** (0.004)	0.995*** (0.004)	0.995*** (0.005)	0.993*** (0.005)
R <sup>2</sup>	0.434	0.527	0.437	0.529	0.957	0.957	0.957	0.957
Linear spline	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic spline	No	No	Yes	Yes	No	No	Yes	Yes
Baseline controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Each column contains the first stage, the estimated effect of the intent to treat on treatment status. See table 1 for additional details on student, school, and teacher controls. Fuzzy regression discontinuity (RD)  $N = 920$ ; Frontier RD  $N = 767$ . Robust standard errors in parentheses. SGG = Student Gap Group.

\*\*\*  $p < 0.001$ .

75 percentage points at the threshold. In the subsequent specifications that introduce additional controls (e.g., quadratic splines and baseline covariates), the point estimate is 68 to 70 percentage points. As expected, in the Frontier RD application, the “jump” in treatment status at the threshold is nearly one, reflecting the virtually sharp first-stage relationship.

These first-stage relationships clearly show a large and discontinuous jump in a school’s treatment status at the arbitrary threshold set by federal policy. Several other ancillary sources of information support our maintained assumption that this RD approach results in credible causal inferences. First, although some schools may feel stigmatized by their Focus status, the multi-tiered assignment protocol based on pre-existing data (i.e., the 2011–12 K-PREP used to generate SGG scores) makes it quite unlikely that schools would have been able to manipulate their treatment assignment. A precise manipulation of the SGG score is also infeasible because it would need to rely on the unduplicated count of students—many of whom could be members of multiple subgroups—that constitute the gap group. Nonetheless, we conduct robustness checks to interrogate this assumption empirically.

For example, following McCrary (2008), we implement a density test to see whether the density of the SGG values jumps discontinuously at the threshold. We fail to reject the null hypotheses that there is no discontinuity at both thresholds for the Fuzzy and Frontier samples (see the online appendix figure A.2). Second, we also examine the density of the assignment variable visually because “heaping” at certain values may fail to be detected by the McCrary density test (Barreca, Lindo, and Waddell 2016). In online appendix figure A.3, we present additional histograms to examine potential heaping-induced bias by inspecting the distribution of the forcing variable with the smallest bin-widths possible. Although there does appear to be some evidence of heaping at several assignment variable values, we follow the suggestions of Barreca, Lindo, and Waddell (2016) for detecting nonrandom heaping and confirmed the robustness of our results.<sup>27</sup>

27. Specifically, we tested for discontinuous jumps in baseline characteristics at data heaps and did not find statistically significant differences. Nonetheless, we also examined our results, omitting observations at possible heaps, and found similar results.

**Table 3.** Auxiliary Regression Discontinuity (RD) Estimates of Baseline Covariate Balance

Dependent Variable	Math Achievement Index		Reading Achievement Index	
	Fuzzy RD	Frontier RD	Fuzzy RD	Frontier RD
Sample				
Gap group	0.102 (0.791)	0.493 (0.962)	-1.523 (0.844)	-0.110 (1.052)
Non-gap group	1.909 (1.231)	0.807 (1.621)	0.080 (0.820)	0.461 (1.122)
Free/reduced-price lunch	-0.316 (0.812)	0.085 (0.928)	-1.661 (0.934)	0.319 (1.121)
Special education	-1.218 (1.660)	-0.592 (2.290)	-2.720 (2.042)	-1.395 (2.876)

Notes: Each cell contains the results of a two-stage regression: (1) in the first stage the 2013–14 outcome for the relevant subgroup is regressed on all baseline covariates (see table 1 for a description of covariates) and a predicted achievement composite is generated. (2) In the second stage the predicted achievement composite (i.e., portion of the proficiency rate predicted by the baseline covariates) is then regressed on  $I(\text{Student Gap Group Score}_i < 0)$  and a linear spline for the assignment variable. Robust standard errors in parentheses.

Third, we also examined the baseline traits of schools listed in table 1 to assess whether they are continuous across the assignment variable threshold. In table 3, we present evidence of this covariate balance based on a simple two-stage process. In the first stage, we regressed the outcome of interest, the 2013–14 proficiency rate in math or reading, on all of the baseline covariates. We then obtain predicted proficiency rates based on these regressions. These predicted values provide a single value for each school that represents an index of all the baseline traits that influence outcomes, each weighted by its regression-estimated impacts. We then estimate and report in table 3 auxiliary RD equations in which these indices are the dependent variables.<sup>28</sup> In a valid RD design, our expectation is that these indices of baseline school traits would be continuous through the threshold that defines treatment status. The results in table 3 consistently indicate, for different estimation methods (and different samples), that the covariates are balanced around the threshold.<sup>29</sup> Overall, these results clearly suggest that the assignment of Kentucky's schools to Focus School reforms functions as a well-behaved regression discontinuity.

## 6. RESULTS

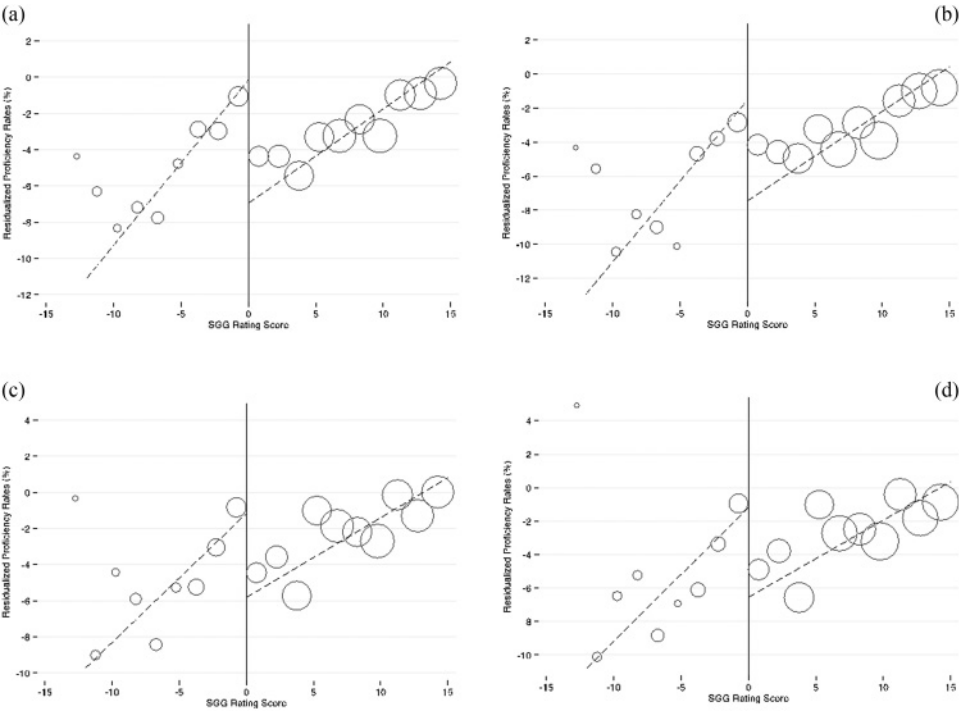
### Main Results

We begin by illustrating our main findings graphically in figure 3. These four panels show the 2013–14 math and reading proficiency rates of gap-group students in schools above and below the treatment-eligibility threshold, and do so both for the full Fuzzy sample and the smaller Frontier sample. These figures consistently indicate that, after the first full year of implementation, the math and reading proficiency of gap-group students jumped meaningfully at the threshold that influenced Focus School status.<sup>30</sup> For

28. The use of these indices attenuates the multiple comparison problems that might plague a plethora of tests based on each baseline covariate. Nonetheless, we report estimates for the available individual covariates in the online appendix (table A.4).

29. We also examine each variable separately, visually and parametrically, and do not find evidence of changes in pretreatment characteristics at the assignment threshold.

30. Although we focus on the first full year of treatment, we also examine academic outcomes in the 2012–13 school year and find positive achievement effects that are smaller in magnitude but statistically insignificant.



Notes: Graphs of 2014 residualized proficiency rate for math and reading by the Student Gap Group assignment variable for Fuzzy and Frontier RD samples. Bin width 1.5, bandwidth  $\pm 15$ , Markers weighted by number of schools in bin width.

**Figure 3.** Gap Group Proficiency Rates, by Regression Discontinuity (RD) Sample, 2013–14. (a) 2013–14 Math Proficiency Residualized, Gap Group, Fuzzy RD. (b) 2013–14 Math Proficiency Residualized, Gap Group, Frontier RD. (c) 2013–14 Reading Proficiency Residualized, Gap Group, Fuzzy RD. (d) 2014 Reading Proficiency Residualized, Gap Group, Frontier RD.

reading, the jumps in both samples appear to be of a slightly smaller magnitude compared with the substantial jump for math. In table 4, we provide parametric evidence that supports these graphical results.

Table 4 reports the estimated reduced-form effect of treatment eligibility on reading and math proficiency across different specifications (i.e., with and without baseline controls and in the Fuzzy and Frontier samples). Panel A of table 4 shows the math results, which consistently indicate impacts of roughly 5 percentage points. Because these are intent-to-treat estimates and the Fuzzy RD has lower first-stage compliance, the “treatment on the treated” effects implied by these results are somewhat larger in the Fuzzy RD specification. To put estimates of this size in perspective, we note that the math proficiency rate for gap-group students in schools just above the eligibility threshold is approximately 30 percent. This implies that a 5-percentage-point increase is equivalent to a 17-percent increase relative to the counterfactual mean.

Panel B of table 4 reports the estimated effect of Focus School eligibility on reading proficiency rates among gap-group students. These intent-to-treat estimates indicate that Focus School eligibility increased reading proficiency by 2.3 to 3.6 percentage

We hypothesize this is due to Focus Schools’ inability to fully implement the planned reforms (although it is impossible to disentangle this effect from those that may be associated with the stigma of the Focus label).

**Table 4.** The Effect of Focus School Eligibility on Gap Group Proficiency Rates by Subject

Panel A: Math Proficiency Rate				
Independent Variable	Fuzzy RD		Frontier RD	
	(1)	(2)	(3)	(4)
$I(\text{SGG Score}_i < 0)$	4.361*** (1.651)	5.166*** (1.641)	4.586** (1.976)	4.754** (1.976)
$R^2$	0.343	0.425	0.304	0.382
$N$	916	916	765	765
Panel B: Reading Proficiency Rate				
Independent Variable	Fuzzy RD		Frontier RD	
	(1)	(2)	(3)	(4)
$I(\text{SGG Score}_i < 0)$	0.373 (1.177)	2.349** (1.088)	2.518* (1.371)	3.555*** (1.332)
$R^2$	0.418	0.487	0.366	0.427
$N$	918	918	765	765
Baseline controls	No	Yes	No	Yes

Notes: Each cell contains the result of a separate regression of the effect of  $I(\text{SGG Score}_i < 0)$  on 2013–14 proficiency rates for gap group students. All models condition on a linear spline of the assignment variable. Akaike's information criterion implied the optimal order of polynomial is linear. Robust standard errors in parentheses. RD = regression discontinuity; SGG = Student Gap Group.

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.10$ .

points.<sup>31</sup> The implied “treatment on the treated” estimates (i.e., the estimated effect of being a Focus School) are between 3 and 3.5 percentage points. Reading proficiency rates are approximately 37 percent for gap-group students in schools just above the assignment threshold. This implies that Focus School reforms increased reading proficiency by roughly 9 percent relative to the counterfactual mean (i.e.,  $3.5/37$ ).

These results indicate that the Focus School reforms led to substantial gains in math and reading proficiency among the broad-gap group. These performance gains are also reflected in subsequent school exits from Focus School status, which is explicitly based on the performance of gap-group students. The KDE used the 2013–14 achievement results to identify (and reclassify) Focus Schools that met performance targets (and to designate new Focus Schools). The treatment effects documented here contributed to the exit of forty-seven Focus Schools from this status.

### Teacher Survey Results

We use data from the teacher surveys described in section 3 to conduct auxiliary RD regressions that examine the effects of Focus School reforms on teachers' perceptions of the quality of their professional development (i.e., ten distinct items and an overall composite).<sup>32</sup> The existing research on the ability of teacher professional development to improve student achievement is overwhelmingly thin yet professional

31. In the Fuzzy RD specification that omits baseline controls, this estimate is smaller and statistically insignificant. However, we examined the AIC for this and all other models, and found the AIC privileges those that condition on both linear splines of the assignment variable and the baseline controls.

32. We run auxiliary regressions to check for continuity of response rate and likelihood of any responses at the Focus School assignment threshold, and find no discontinuous jumps, indicating that Focus Schools and control schools had similar response rates at the assignment threshold.



**Table 5.** The Effect of Focus School Eligibility on Proficiency Rates, by Alternative Bandwidths

Bandwidth Sample	Math Proficiency Rate, Gap Group				Reading Proficiency Rate, Gap Group			
	Fuzzy RD		Frontier RD		Fuzzy RD		Frontier RD	
	(1) Estimate	(2) n (Focus)	(3) Estimate	(4) n (Focus)	(5) Estimate	(6) n (Focus)	(7) Estimate	(8) n (Focus)
Full sample	<b>5.166***</b> (1.641)	916 (187)	<b>4.754**</b> (1.976)	765 (49)	<b>2.349**</b> (1.088)	918 (187)	<b>3.555***</b> (1.332)	765 (49)
SGG Score <sub>i</sub>   ≤ 20	4.854*** (1.707)	715 (182)	4.132** (2.066)	570 (49)	1.819 (1.149)	717 (182)	2.823** (1.418)	570 (49)
SGG Score <sub>i</sub>   ≤ 15	4.135** (1.730)	574 (175)	3.287 (2.152)	435 (48)	1.513 (1.270)	576 (175)	2.239 (1.617)	435 (48)
SGG Score <sub>i</sub>   ≤ 10	4.512** (1.984)	387 (142)	3.477 (2.534)	281 (43)	2.443 (1.513)	388 (142)	3.545* (1.893)	281 (43)
SGG Score <sub>i</sub>   ≤ 7	4.501* (2.343)	252 (113)	2.380 (2.854)	170 (36)	4.084** (1.775)	253 (113)	4.962** (2.176)	170 (36)
Kernel regression	4.727** (2.222)	368 (141)	3.183 (2.669)	264 (40)	3.917** (1.585)	369 (138)	5.028*** (1.823)	264 (40)

Notes: **Bold** text represents the main results that are presented in table 4. Each cell contains a regression of 2013–14 reading or math proficiency rates for schools within the specified bandwidth on  $I(\text{SGG Score}_i < 0)$ , a linear spline of the assignment variable and baseline covariates. Columns 2, 4, 6, and 8 contain the number of schools in the specified bandwidth and the number of treatment schools below the bandwidth total in parentheses. See table 1 for a description of controls. Robust standard errors in parentheses. RD = regression discontinuity; SGG = Student Gap Group.

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.10$ .

development continues to be a cornerstone of many school reform strategies (Yoon et al. 2007; Gersten et al. 2014). In online table A.3, we present results from the survey that are uniformly positive, indicating that Focus School reforms led to improvements in the quality of teacher professional development. The gains were particularly large (and statistically significant) for teacher reports of whether their professional development included collaboration and follow-up, as well as whether it supported teachers in meeting student needs and promoting learning. Although we are unable to definitively link the quality of professional development to the positive outcomes for students at Focus Schools, we provide this evidence as a potential mechanism given Kentucky's waiver application, which emphasized high-quality professional development.

### Robustness Checks

We perform a number of robustness checks to test the sensitivity of our main results. First, we examine the robustness of our results to functional-form concerns by reporting the results from local linear regressions. In these local linear regression specifications, we condition on a linear spline of the assignment variable and the baseline controls but limit the sample to increasingly tight bandwidths of schools proximate to the eligibility threshold. These bandwidth restrictions include the optimal bandwidths (i.e.,  $\pm 7$  to 9 assignment score points) based on the procedure introduced by Imbens and Kalyanaraman (2011). We also report weighted least squares estimates based on a triangular kernel weight that takes higher values for schools close to either side of the eligibility threshold.

Table 5 reports the key results from these specifications for both math and reading results as well as for the Fuzzy and Frontier RD samples. The results across these

different bandwidths and outcomes consistently indicate that the Focus School reforms had qualitatively large and positive effects on student proficiency in math and reading. In most cases, the estimates based on these smaller samples also remain statistically significant. However, we note two broad exceptions. First, in the smaller Frontier RD sample, further bandwidth restrictions imply that the estimated effects on math, though still positive, become smaller and substantially noisier (i.e., estimates in column 3). Second, with regard to the reading results, modest bandwidth restrictions (i.e., within 10 to 20 points of the threshold) imply estimated effects that are smaller and statistically insignificant. However, further bandwidth restrictions and the kernel-weighted results imply consistent and statistically significant effects on reading proficiency that are qualitatively similar to those based on the full samples.

As an additional robustness check, we present several auxiliary RD estimates at “placebo” RD thresholds (see online appendix table A.5). We artificially move the treatment threshold to various placebo cut scores to examine whether the positive treatment effects are evident at other unimportant thresholds that hold no policy significance for schools. This additional check allows us to increase confidence in our main result estimates, which are significant, while the placebo thresholds are not. None of the placebo estimates are consistently statistically significant and further bolster our main results on the effects of Kentucky’s Focus School reforms.

Finally, because our analyses leverage school-level data, we may be concerned that changes in the composition of students enrolled at schools (or those taking the test) drives the positive results. To examine this concern, we first examined auxiliary RD specifications in which the number of students tested (i.e., in the gap group as well as other subgroups) was the dependent variable. These results consistently indicated no change in the size of these test-taking populations at the eligibility threshold. We adopted a similar approach to examine whether nonrandom mobility into or out of the school may be a source of bias. First, we estimated a regression of the 2013–14 performance outcomes on a variety of post-treatment school covariates reflecting the size and composition of the school (e.g., total enrollment, percent racial/ethnic enrollment, proportion qualifying for free or reduced-priced lunch, LEP, and special education participation). We then took the predicted values from these regressions and used them as the dependent variables in auxiliary RD specifications. This approach allows us to examine whether an index of these post-treatment school traits, weighted by their relevance for student outcomes, differs above and below the eligibility threshold. The key results reported in table 6 indicate these post-treatment school traits do not differ significantly around the threshold.<sup>33</sup> This finding is consistent with the absence of treatment-endogenous student sorting and, by implication, the causal warrant of the RD results.

### Treatment Heterogeneity

Another relevant concern with respect to our main findings is that they may be misleading if the state’s focus on proficiency encouraged treatment schools to focus their reforms largely on students close to the proficiency threshold (Lazear 2006; Neal and Schanzenbach 2010). In table 7, we present evidence that speaks to this concern by estimating the effects of Focus School eligibility on alternative test thresholds. Moving

33. We also examine these post-treatment characteristics individually and find no evidence of nonrandom sorting.

**Table 6.** Auxiliary Regression Discontinuity (RD) Estimates of Post-treatment Covariate Balance

	Fuzzy RD	Frontier RD
Math achievement index	0.269 (0.456)	0.0328 (0.624)
Reading achievement index	0.344 (0.414)	0.138 (0.478)

*Notes:* Each cell contains the results of a two-stage regression: (1) in the first stage the 2013–14 test score for the gap subgroup is regressed on all post-treatment school covariates (i.e., percent racial/ethnic enrollment, total enrollment, attendance, student–teacher ratio, percent free/reduced-price lunch, percent limited English proficient, percent special education, and number of tested students in a subgroup) and a predicted composite is generated. (2) In the second stage the predicted post-treatment composite is regressed on  $I(\text{Student Gap Group Score}_i < 0)$  and a linear spline for the assignment variable and baseline covariates (see table 1 for a description of covariates). Robust standard errors in parentheses.

**Table 7.** The Effect of Focus School Eligibility on Gap Group Proficiency Levels

Math			
Performance Level	Fuzzy RD Estimate	Frontier RD Estimate	Mean
Apprentice and above	2.828* (1.591)	2.392 (1.841)	78.5
Proficient and above	<b>5.166***</b> (1.641)	<b>4.754**</b> (1.976)	38.8
Distinguished and above	0.993 (0.678)	0.430 (0.938)	8.6
<i>N</i>	916	765	
Reading			
Apprentice and above	2.393* (1.246)	3.484** (1.522)	74.3
Proficient and above	<b>2.349**</b> (1.088)	<b>3.555***</b> (1.332)	46.2
Distinguished and above	1.780** (0.719)	2.172** (0.981)	12.5
<i>N</i>	918	765	

*Notes:* **Bold** text represents the main results that are presented in table 4. Each cell contains a separate regression of the effect of  $I(\text{Student Gap Group Score}_i < 0)$  on 2013–14 proficiency levels for gap group students. All models contain a linear spline for the assignment variable and controls from the preferred specification in table 4. The lowest proficiency level is novice followed by apprentice, proficient, and distinguished. Robust standard errors in parentheses. RD = regression discontinuity.

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.10$ .

a student from the lowest performance level to the apprentice level would not enable a school to exit Focus School status. Similarly, moving a proficient student to distinguished status also does not influence Focus School status. Therefore, if schools were adopting strategic “triage” behavior, we would not expect to see treatment effects at these margins. However, the RD results in table 7 indicate that Focus School eligibility generated positive and, in most cases, statistically significant effects at each performance threshold. One notable exception is that the estimated effects on distinguished

**Table 8.** The Effect of Focus School Eligibility on Subgroup Proficiency Rates

Sample	Math Proficiency Rate				Reading Proficiency Rate			
	Fuzzy RD	N	Frontier RD	N	Fuzzy RD	N	Frontier RD	N
Subgroup								
Gap group	<b>5.166***</b> (1.641)	916	<b>4.754**</b> (1.976)	765	<b>2.349**</b> (1.088)	918	<b>3.555***</b> (1.332)	765
Free/reduced-price lunch	<b>4.077**</b> (1.614)	905	3.128 (2.081)	754	2.003* (1.186)	904	3.033* (1.441)	752
Special education	<b>6.378**</b> (2.578)	854	5.294 (3.783)	707	2.613 (2.422)	852	4.293 (3.309)	706
Non-gap	3.365 (3.167)	914	<b>8.531**</b> (4.183)	764	1.528 (2.941)	916	7.272* (3.272)	764

Notes: **Bold** text represents the main results that are presented in table 4. Each cell contains the result of a separate regression of  $I(\text{Student Gap Group Score} < 0)$  on proficiency rates for different subgroups. All models condition on a linear spline of the assignment variable and the full set of controls from the preferred specification in table 4. Robust standard errors in parentheses. RD = regression discontinuity.

\*\*\*  $p < 0.01$ ; \*\*  $p < 0.05$ ; \*  $p < 0.10$ .

performance in mathematics are much smaller, possibly reflecting the comparative rarity of performance at this level (i.e., 8.6 percent in this sample).

Our analysis so far has focused on the test performance of gap-group students, an amalgamation of traditionally low-performing subgroups and the focal point of these reforms. However, the effects of these reforms on other groups of students are also of interest. For example, it may be that these reforms, which had a schoolwide character (e.g., improvement planning, teacher professional development), had spillover benefits for students outside the gap group.<sup>34</sup> Alternatively, it may be that schools strategically reallocated instructional effort away from such “non-gap” students. Although Kentucky does not specifically report on the performance of non-gap group students, we are able to construct performance level percentages from available information on the performance of all students and gap-group students. We also have school performance data on subgroups of students qualifying for free or reduced-price lunch and students receiving special education services (i.e., two categories in the gap group). In table 8, we present RD estimates of the effect of Focus School eligibility on the proficiency rates of these student groups. The results are consistently positive, though not always statistically significant. These results suggest that the performance benefits of the Focus School reforms were broad and, in particular, did not harm (and may have helped) the performance of students outside the gap group.

## 7. CONCLUSION

The conduct and performance of public schools in the United States have historically been the concern of local and state authorities. However, over the last fifteen years, the federal government has been unusually prescriptive in seeking to improve the performance of public schools, particularly with regard to historically marginalized students. The most recent iteration of these controversial, federally catalyzed reforms

34. It should be noted that gap-group students constitute the majority of enrollment in the average Kentucky school (e.g., the average school has 63 percent of students qualifying for free or reduced-price lunch, all of whom would be part of the gap group).

occurred under waivers to NCLB. A signature feature of these waiver reforms was the push for “differentiated accountability.” In order to address achievement gaps, these waivers required states to identify schools where targeted subgroups performed particularly poorly and to support reforms in these designated Focus Schools. In this study, we examined the causal effects of the Focus School reforms in the state of Kentucky. Our RD results indicate that these reforms led to substantial improvements in the math and reading performance of the targeted students (and, possibly, among the students who were not targeted as well). We also present evidence that comprehensive school planning, including an emphasis on high-quality teacher professional development, were mediators of these effects.

The effectiveness of Focus School reforms in Kentucky stands in notable contrast to what we have learned about similar reform efforts in other states (Dee and Dizon-Ross 2017; Dougherty and Weiner 2017; Hemelt and Jacob 2017) where waiver-driven reforms appear to have had no or at best targeted effects. One ready explanation for the comparative success of Focus School reforms in Kentucky involves the state-specific context. The state of Kentucky has established a reputation as an enthusiastic and energetic adopter and implementer of school reforms. The Focus School reforms in Kentucky appeared to reflect this. The federal requirements for Focus School reforms were vague but Kentucky articulated detailed and comprehensive reform activities for Focus Schools to undertake. This engagement and the articulation of “home grown” reforms are both likely to complement a high-fidelity implementation. Furthermore, the Focus School reforms in Kentucky also had a fairly unique design feature: the designation of a large umbrella group of underserved students for which schools would be held responsible. The use of this larger gap group may have been instrumental in catalyzing broader schoolwide reforms and avoiding incentives for narrowly targeted reform efforts. Learning more about how state and local contexts may channel reform efforts productively is likely to be an increasingly important area of study in education policy.

## ACKNOWLEDGMENTS

We would like to acknowledge financial support from the Spencer, Walton, and William T. Grant Foundations, as well as the Institute of Education Sciences (grant R305B140009). We also express appreciation for comments provided by seminar participants at Stanford University and by participants at the AEFPP and APPAM research conferences.

## REFERENCES

- Barreca, Alan I., Jason M. Lindo, and Glen R. Waddell. 2016. Heaping-induced bias in regression-discontinuity designs. *Economic Inquiry* 54(1): 268–293.
- Bifulco, Robert, William Duncombe, and John Yinger. 2005. Does whole-school reform boost student performance? The case of New York City. *Journal of Policy Analysis and Management* 24(1): 47–72.
- Borman, Geoffrey D., Gina M. Hewes, Laura T. Overman, and Shelly Brown. 2003. Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research* 73(2): 125–230.
- Carnoy, Martin, and Susanna Loeb. 2002. Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis* 24(4): 305–331.

- Cook, Thomas D., John Deke, Lisa Dragoset, Sean F. Reardon, Rocio Titunik, Petra Todd, Wilbert van der Klaauw, and Glenn Wadell. 2015. *Preview of regression discontinuity design standards*. Available [https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc\\_rdd\\_standards\\_122315.pdf](https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_rdd_standards_122315.pdf). Accessed 20 August 2019.
- Dee, Thomas S. 2012. School turnarounds: Evidence from the 2009 stimulus. NBER Working Paper No. 17990.
- Dee, Thomas S., and Elise Dizon-Ross. 2017. School performance, accountability and waiver reforms: Evidence from Louisiana. NBER Working Paper No. 23463.
- Dee, Thomas S., and Brian Jacob. 2011. The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management* 30(3): 418–446.
- Dee, Thomas S., Brian Jacob, and Nathaniel L. Schwartz. 2013. The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis* 35(2): 252–279.
- Delisle, Deborah S. 2014. *Letter to Commissioner Terry Holliday*. Available [www2.ed.gov/policy/eseaflex/secretary-letters/ky2extltr814.pdf](http://www2.ed.gov/policy/eseaflex/secretary-letters/ky2extltr814.pdf). Accessed 20 August 2019.
- Desimone, Laura. 2002. How can comprehensive school reform models be successfully implemented? *Review of Educational Research* 72(3): 433–479.
- Dougherty, Shaun M., and Jennie M. Weiner. 2017. The Rhode to turnaround: The impact of waivers to No Child Left Behind on school performance. *Educational Policy* 33(4): 555–586.
- Figlio, David, and Susanna Loeb. 2011. School accountability. In *Handbook of the economics of education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessman, pp. 383–417. Amsterdam: North Holland.
- Gersten, Russell, Mary Jo Taylor, Tran D. Keys, Eric Rolfhus, and Rebecca Newman-Gonchar. 2014. *Summary of research on the effectiveness of math professional development approaches*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.
- Gewertz, Catherine. 2011. Kentucky to model professional development for new standards. *Education Week*, 4 November.
- Governor's Task Force. 2011. *Breaking new ground: Final report of the Governor's Task Force on transforming education in Kentucky*. Frankfort: Kentucky Department of Education.
- Gross, Betheny, Kevin T. Booker, and Dan Goldhaber. 2009. Boosting student achievement: The effect of comprehensive school reform on student achievement. *Educational Evaluation and Policy Analysis* 31(2): 111–126.
- Hemelt, Steven W., and Brian Jacob. 2017. Differentiated accountability and education production: Evidence from NCLB Waivers. NBER Working Paper No. 23461.
- Imbens, Guido, and Karthik Kalyanaraman. 2011. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3): 933–959.
- Kentucky Department of Education (KDE). 2011. *Guidelines for closing the gaps for all students*. Available <https://education.uky.edu/wp-content/uploads/2018/10/Guidelines-for-Closing-the-Gap-for-All-Students.pdf>. Accessed 20 August 2019.
- Klein, Alyson. 2015. Arne Duncan stepping down as Education Secretary: Hefty policy footprint over tumultuous term. *Education Week*, 10 October.
- Klein, Alyson. 2016. The Every Student Succeeds Act: An ESSA overview. *Education Week*, 15 March.

- Klein, Alyson. 2017. ESSA plans: Takeaways from the first batch of approvals. *Education Week*, 20 September.
- Lazear, Edward P. 2006. Speeding, terrorism, and teaching to the test. *Quarterly Journal of Economics* 121(3): 1029–1061.
- Lee, David S., and Thomas Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2): 281–355.
- McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2): 698–714.
- Neal, Derek, and Diane Whitmore Schanzenbach. 2010. Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics* 92(2): 263–283.
- Papay, John, and Molly Hannon. 2015. The effects of school turnaround strategies in Massachusetts. Paper presented at Association of Public Policy and Management Conference, Miami, November.
- Polikoff, Morgan S., Andrew J. McEachin, Stephani L. Wrabel, and Matthew Duque. 2014. The waive of the future? School accountability in the waiver era. *Educational Researcher* 43(1): 45–54.
- Reardon, Sean F., and Joseph P. Robinson. 2012. Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness* 5(1): 83–104.
- Spears, Valarie Honeycutt. 2015. William Wells Brown Elementary struggles to move beyond label as state's lowest performer. *Lexington Herald Leader*, 25 May.
- Sun, Min, Emily K. Penner, and Susanna Loeb. 2017. Resource-and approach-driven multidimensional change: Three-year effects of School Improvement Grants. *American Educational Research Journal* 54(4): 607–643.
- Ujifusa, Andrew. 2013. Kentucky Chief Holliday is new CCSSO president; Will focus on career readiness. *Education Week*, 31 December.
- U.S. Department of Education (USDOE). 2010. Evaluation of the comprehensive school reform program implementation and outcomes: Fifth-year report. Washington, DC: USDOE, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.
- U.S. Department of Education (USDOE). 2011. *Kentucky ESEA flexibility request: Final submission*. Available [www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ky.html](http://www2.ed.gov/policy/elsec/guid/esea-flexibility/map/ky.html). Accessed 20 August 2019.
- U.S. Department of Education (USDOE). 2012. *ESEA flexibility*. Available [www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html](http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html). Accessed 20 August 2019.
- U.S. Department of Education (USDOE). 2013. *ESEA flexibility part B monitoring report*. Available <https://www2.ed.gov/admins/lead/account/monitoring/reports13/kypartbrpt2014.pdf>. Accessed 20 August 2019.
- Wong, Vivian C., Peter M. Steiner, and Thomas D. Cook. 2013. Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics* 38(2): 107–141.
- Yoon, Kwang S., Teresa Duncan, Silvia W. Lee, Beth Scarloss, and Kathy L. Shapley. 2007. *Reviewing the evidence on how teacher professional development affects student achievement*. Available [https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL\\_2007033.pdf](https://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf). Accessed 20 August 2019.